

# Enhancing the Quality and Credibility of Qualitative Analysis

*Michael Quinn Patton*

---

**Abstract.** Varying philosophical and theoretical orientations to qualitative inquiry remind us that issues of quality and credibility intersect with audience and intended research purposes. This overview examines ways of enhancing the quality and credibility of qualitative analysis by dealing with three distinct but related inquiry concerns: rigorous techniques and methods for gathering and analyzing qualitative data, including attention to validity, reliability, and triangulation; the credibility, competence, and perceived trustworthiness of the qualitative researcher; and the philosophical beliefs of evaluation users about such paradigm-based preferences as objectivity versus subjectivity, truth versus perspective, and generalizations versus extrapolations. Although this overview examines some general approaches to issues of credibility and data quality in qualitative analysis, it is important to acknowledge that particular philosophical underpinnings, specific paradigms, and special purposes for qualitative inquiry will typically include additional or substitute criteria for assuring and judging quality, validity, and credibility. Moreover, the context for these considerations has evolved. In early literature on evaluation methods the debate between qualitative and quantitative methodologists was often strident. In recent years the debate has softened. A consensus has gradually emerged that the important challenge is to match appropriately the methods to empirical questions and issues, and not to universally advocate any single methodological approach for all problems.

**Key Words.** Qualitative research, credibility, research quality

---

Approaches to qualitative inquiry have become highly diverse, including work informed by phenomenology, ethnomethodology, ethnography, hermeneutics, symbolic interaction, heuristics, critical theory, positivism, and feminist inquiry, to name but a few (Patton 1990). This variety of philosophical and theoretical orientations reminds us that issues of quality and credibility intersect with audience and intended research purposes. Research directed to an audience of independent feminist scholars, for example, may be judged by criteria somewhat different from those of research addressed to an audience of

government policy researchers. Formative research or action inquiry for program improvement involves different purposes and therefore different criteria of quality compared to summative evaluation aimed at making fundamental continuation decisions about a program or policy (Patton 1997). New forms of personal ethnography have emerged directed at general audiences (e.g., Patton 1999). Thus, while this overview will examine some general qualitative approaches to issues of credibility and data quality, it is important to acknowledge that particular philosophical underpinnings, specific paradigms, and special purposes for qualitative inquiry will typically include additional or substitute criteria for assuring and judging quality, validity, and credibility.

## THE CREDIBILITY ISSUE

The credibility issue for qualitative inquiry depends on three distinct but related inquiry elements:

- rigorous techniques and methods for gathering high-quality data that are carefully analyzed, with attention to issues of validity, reliability, and triangulation;
- the credibility of the researcher, which is dependent on training, experience, track record, status, and presentation of self; and
- philosophical belief in the value of qualitative inquiry, that is, a fundamental appreciation of naturalistic inquiry, qualitative methods, inductive analysis, purposeful sampling, and holistic thinking.

## TECHNIQUES FOR ENHANCING THE QUALITY OF ANALYSIS

Although many rigorous techniques exist for increasing the quality of data collected, at the heart of much controversy about qualitative findings are doubts about the nature of the analysis. Statistical analysis follows formulas and rules while, at the core, qualitative analysis is a creative process, depending on the insights and conceptual capabilities of the analyst. While creative approaches may extend more routine kinds of statistical analysis,

---

Address correspondence to Michael Quinn Patton, Professor, Utilization-Focused Evaluation, The Union Institute, 3228 46th Ave. South, Minneapolis, MN 55406. This article, submitted to *Health Services Research* on January 13, 1999, was revised and accepted for publication on June 24, 1999.

qualitative analysis depends from the beginning on astute pattern recognition, a process epitomized in health research by the scientist working on one problem who suddenly notices a pattern related to a quite different problem. This is not a simple matter of chance, for as Pasteur posited: "Chance favors the prepared mind."

More generally, beyond the analyst's preparation and creative insight, there is also a technical side to analysis that is analytically rigorous, mentally replicable, and explicitly systematic. It need not be antithetical to the creative aspects of qualitative analysis to address issues of validity and reliability. The qualitative researcher has an obligation to be methodical in reporting sufficient details of data collection and the processes of analysis to permit others to judge the quality of the resulting product.

### *Integrity in Analysis: Testing Rival Explanations*

Once the researcher has described the patterns, linkages, and plausible explanations through inductive analysis, it is important to look for rival or competing themes and explanations. This can be done both inductively and logically. Inductively it involves looking for other ways of organizing the data that might lead to different findings. Logically it means thinking about other logical possibilities and then seeing if those possibilities can be supported by the data. When considering rival organizing schemes and competing explanations the mind-set is not one of attempting to disprove the alternatives; rather, the analyst looks for data that support alternative explanations. Failure to find strong supporting evidence for alternative ways of presenting the data or contrary explanations helps increase confidence in the original, principal explanation generated by the analyst. It is likely that comparing alternative explanations or looking for data in support of alternative patterns will not lead to clear-cut "yes, there is support" versus "no, there is not support" kinds of conclusions. It is a matter of considering the weight of evidence and looking for the best fit between data and analysis. It is important to report what alternative classification systems, themes, and explanations are considered and "tested" during data analysis. This demonstrates intellectual integrity and lends considerable credibility to the final set of findings offered by the evaluator.

### *Negative Cases*

Closely related to the testing of alternative constructs is the search for negative cases. Where patterns and trends have been identified, our understanding of

those patterns and trends is increased by considering the instances and cases that do not fit within the pattern. These may be exceptions that prove the rule. They may also broaden the "rule," change the "rule," or cast doubt on the "rule" altogether. For example, in a health education program for teenage mothers where the large majority of participants complete the program and show knowledge gains, an important consideration in the analysis may also be examination of reactions from drop-outs, even if the sample is small for the drop-out group. While perhaps not large enough to make a statistical difference on the overall results, these reactions may provide critical information about a niche group or specific subculture, and/or clues to program improvement.

I would also note that the section of a qualitative report that involves an exploration of alternative explanations and a consideration of why certain cases do not fall into the main pattern can be among the most interesting sections of a report to read. When well written, this section of a report reads something like a detective study in which the analyst (detective) looks for clues that lead in different directions and tries to sort out the direction that makes the most sense given the clues (data) that are available.

### *Triangulation*

The term "triangulation" is taken from land surveying. Knowing a single landmark only locates you somewhere along a line in a direction from the landmark, whereas with two landmarks you can take bearings in two directions and locate yourself at their intersection. The notion of triangulating also works metaphorically to call to mind the world's strongest geometric shape—the triangle (e.g., the form used to construct geodesic domes à la Buckminster Fuller). The logic of triangulation is based on the premise that no single method ever adequately solves the problem of rival explanations. Because each method reveals different aspects of empirical reality, multiple methods of data collection and analysis provide more grist for the research mill.

Triangulation is ideal, but it can also be very expensive. A researcher's limited budget, short time frame, and narrow training will affect the amount of triangulation that is practical. Combinations of interview, observation, and document analysis are expected in much fieldwork. Studies that use only one method are more vulnerable to errors linked to that particular method (e.g., loaded interview questions, biased or untrue responses) than are studies that use multiple methods in which different types of data provide cross-data validity checks.

It is possible to achieve triangulation within a qualitative inquiry strategy by combining different kinds of qualitative methods, mixing purposeful samples, and including multiple perspectives. It is also possible to cut across inquiry approaches and achieve triangulation by combining qualitative and quantitative methods, a strategy discussed and illustrated in the next section. Four kinds of triangulation contribute to verification and validation of qualitative analysis: (1) checking out the consistency of findings generated by different data collection methods, that is, *methods triangulation*; (2) examining the consistency of different data sources within the same method, that is, *triangulation of sources*; (3) using multiple analysts to review findings, that is, *analyst triangulation*; and (4) using multiple perspectives or theories to interpret the data, that is, *theory/perspective triangulation*. By combining multiple observers, theories, methods, and data sources, researchers can make substantial strides in overcoming the skepticism that greets singular methods, lone analysts, and single-perspective theories or models.

However, a common misunderstanding about triangulation is that the point is to demonstrate that different data sources or inquiry approaches yield essentially the same result. But the point is really to test for such consistency. Different kinds of data may yield somewhat different results because different types of inquiry are sensitive to different real world nuances. Thus, an understanding of inconsistencies in findings across different kinds of data can be illuminative. Finding such inconsistencies ought not be viewed as weakening the credibility of results, but rather as offering opportunities for deeper insight into the relationship between inquiry approach and the phenomenon under study.

### *Reconciling Qualitative and Quantitative Data*

Methods triangulation often involves comparing data collected through some kinds of qualitative methods with data collected through some kinds of quantitative methods. This is seldom straightforward because certain kinds of questions lend themselves to qualitative methods (e.g., developing hypotheses or a theory in the early stages of an inquiry, understanding particular cases in depth and detail, getting at meanings in context, capturing changes in a dynamic environment), while other kinds of questions lend themselves to quantitative approaches (e.g., generalizing from a sample to a population, testing hypotheses, or making systematic comparisons on standardized criteria). Thus, it is common that quantitative methods and qualitative methods are used in a complementary fashion to answer different questions that do not easily come together to provide a single, well-integrated picture of the

situation. Moreover, few researchers are equally comfortable with both types of data, and the procedures for using the two together are still emerging. The tendency is to relegate one type of analysis or the other to a secondary role according to the nature of the research and the predilections of the investigators. For example, observational data are often used for generating hypotheses or describing processes, while quantitative data are used to make systematic comparisons and verify hypotheses. Although it is common, such a division of labor is also unnecessarily rigid and limiting.

Given the varying strengths and weaknesses of qualitative versus quantitative approaches, the researcher using different methods to investigate the same phenomenon should not expect that the findings generated by those different methods will automatically come together to produce some nicely integrated whole. Indeed, the evidence is that one ought to expect that initial conflicts will occur between findings from qualitative and quantitative data and that those findings will be received with varying degrees of credibility. It is important, then, to consider carefully what each kind of analysis yields and to give different interpretations the chance to arise and be considered on their merits before favoring one result over another based on methodological biases.

In essence, triangulation of qualitative and quantitative data is a form of comparative analysis. "Comparative research," according to Fielding and Fielding (1986: 13), "often involves different operational measures of the 'same' concept, and it is an acknowledgement of the numerous problems of 'translation' that it is conventional to treat each such measure as a separate variable. This does not defeat comparison, but can strengthen its reliability." Subsequently, deciding whether results have converged remains a delicate exercise subject to both disciplined and creative interpretation. Focusing on what is learned by *the degree of convergence* rather than forcing a dichotomous choice—the different kinds of data do or not converge—typically yields a more balanced overall result.

That said, it is worth noting that qualitative and quantitative data can be fruitfully combined when they elucidate complementary aspects of the same phenomenon. For example, a community health indicator (e.g., teenage pregnancy rate) can provide a general and generalizable picture, while case studies of a few pregnant teenagers can put faces on the numbers and illuminate the stories behind the quantitative data; this becomes even more powerful when the indicator is broken into categories (e.g., those under age 15, those 16 and over) with case studies illustrating the implications of and rationale for such categorization.

### *Triangulation of Qualitative Data Sources*

The second type of triangulation involves triangulating data sources. This means comparing and cross-checking the consistency of information derived at different times and by different means within qualitative methods. It means (1) comparing observational data with interview data; (2) comparing what people say in public with what they say in private; (3) checking for the consistency of what people say about the same thing over time; and (4) comparing the perspectives of people from different points of view—for example, in a health program, triangulating staff views, client views, funder views, and views expressed by people outside the program. It means validating information obtained through interviews by checking program documents and other written evidence that can corroborate what interview respondents report. Triangulating historical analyses, life history interviews, and ethnographic participant observations can increase confidence greatly in the final results.

As with triangulation of methods, triangulation of data sources within qualitative methods will seldom lead to a single, totally consistent picture. The point is to study and understand when and why there are differences. The fact that observational data produce different results than those of interview data does not mean that either or both kinds of data are invalid, although that may be the case. More likely, it means that different kinds of data have captured different things and so the analyst attempts to understand the reasons for the differences. At the same time, consistency in overall patterns of data from different sources, and reasonable explanations for differences in data from divergent sources, contribute significantly to the overall credibility of findings.

### *Triangulation Through Multiple Analysts*

The third kind of triangulation is investigator or analyst triangulation, that is, using multiple as opposed to singular observers or analysts. Triangulating analytical processes with multiple analysts is akin to using several interviewers or observers during fieldwork to reduce the potential bias that comes from a single person doing all the data collection, and it provides means to assess more directly the reliability and validity of the data obtained. Having two or more researchers independently analyze the same qualitative data set and then compare their findings provides an important check on selective perception and blind interpretive bias. Another common approach to analytical triangulation is to have those who were studied review the findings. Researchers and evaluators can learn a great deal about the accuracy, fairness,

and validity of their data analysis by having the people described in that data analysis react to what is described. To the extent that participants in the study are unable to relate to the description and analysis in a qualitative evaluation report, it is appropriate to question the credibility of the report. Intended users of findings, especially users of evaluations or policy analyses, provide yet another layer of analytical and interpretive triangulation. In seriously soliciting users' reactions to the face validity of descriptive data, case studies, and interpretations, the evaluator's perspective is joined to the perspective of the people who must use the information. House (1977) suggests that the more "naturalistic" the evaluation, the more it relies on its audiences to reach their own conclusions, draw their own generalizations, and make their own interpretations. House is articulate and insightful on this critical point:

Unless an evaluation provides an explanation for a particular audience, and enhances the understanding of that audience by the content and form of the argument it presents, it is not an adequate evaluation for that audience, even though the facts on which it is based are verifiable by other procedures. One indicator of the explanatory power of evaluation data is the degree to which the audience is persuaded. Hence, an evaluation may be "true" in the conventional sense but not persuasive to a particular audience for whom it does not serve as an explanation. In the fullest sense, then, an evaluation is dependent both on the person who makes the evaluative statement and on the person who receives it (p. 42).

### *Theory Triangulation*

A fourth kind of triangulation involves using different theoretical perspectives to look at the same data. A number of general theoretical frameworks derive from divergent intellectual and disciplinary traditions, e.g., ethnography, symbolic interaction, hermeneutics, or phenomenology (Patton 1990: ch. 3). More concretely, there are always multiple theoretical perspectives that can be brought to bear on substantive issues. One might examine interviews with therapy clients from different psychological perspectives: psychoanalytic, Gestalt, Adlerian, rational-emotive, and/or behavioral frameworks. Observations of a group, community, or organization can be examined from a Marxian or Weberian perspective, a conflict or functionalist point of view. The point of theory triangulation is to understand how findings are affected by different assumptions and fundamental premises.

A concrete version of theory triangulation for evaluation is to examine the data from the perspective of various stakeholder positions with different theories of action about a program. It is common for divergent stakeholders to disagree about program purposes, goals, and means of attaining goals.



These differences represent different theories of action that cast findings in a different light (Patton 1997).

### *Thoughtful, Systematic Triangulation*

These different types of triangulation—methods triangulation, triangulation of data sources, investigator or analyst triangulation, and theory triangulation—are all strategies for reducing systematic bias in the data. In each case the strategy involves checking findings against other sources and perspectives. Triangulation is a process by which the researcher can guard against the accusation that a study's findings are simply an artifact of a single method, a single source, or a single investigator's biases.

### *Design Checks: Keeping Methods and Data in Context*

One primary source of misunderstanding in communicating qualitative findings is over-generalizing the results. By their nature, qualitative findings are highly context and case dependent. Three kinds of sampling limitations typically arise in qualitative research designs:

- Limitations will arise in the situations (critical events or cases) that are sampled for observation (because it is rarely possible to observe all situations).
- Limitations will result from the time periods during which observations took place, that is, problems of temporal sampling.
- Findings will be limited based on selectivity in the people who were sampled either for observations or interviews, or on selectivity in document sampling. This is inherent with purposeful sampling strategies (in contrast to probabilistic sampling).

In reporting how purposeful sampling strategies affect findings, the analyst returns to the reasons for having made initial design decisions. Purposeful sampling involves studying information-rich cases in depth and detail. The focus is on understanding and illuminating important cases rather than on generalizing from a sample to a population. A variety of purposeful sampling strategies are available, each with different implications for the kinds of findings that will be generated (Patton 1990). Rigor in case selection involves explicitly and thoughtfully picking cases that are congruent with the study purpose and that will yield data on major study questions. This sometimes involves a two-stage sampling process where some initial fieldwork is done on a possible case to determine its suitability prior to a commitment to more in-depth and sustained fieldwork.

Examples of purposeful sampling illustrate this point. For instance, sampling and studying highly successful and unsuccessful cases in an intervention yield quite different results than studying a "typical" case or a mix of cases. Findings, then, must be carefully limited, when reported to those situations and cases in the sample. The problem is not one of dealing with a distorted or biased sample, but rather one of clearly delineating the purpose and limitations of the sample studied—and therefore being extremely careful about extrapolating (much less generalizing) the findings to other situations, other time periods, and other people. The importance of reporting both methods and results in their proper contexts cannot be overemphasized and will avoid many controversies that result from yielding to the temptation to generalize from purposeful samples. Keeping findings in context is a cardinal principle of qualitative analysis.

### *The Credibility of the Researcher*

The previous sections have reviewed techniques of analysis that can enhance the quality and validity of qualitative data: searching for rival explanations, explaining negative cases, triangulation, and keeping data in context. Technical rigor in analysis is a major factor in the credibility of qualitative findings. This section now takes up the issue of the effects of the researcher's credibility on the way findings are received.

Because the researcher is the instrument in qualitative inquiry, a qualitative report must include information about the researcher. What experience, training, and perspective does the researcher bring to the field? What personal connections does the researcher have to the people, program, or topic studied? (It makes a difference to know that the observer of an Alcoholics Anonymous program is a recovering alcoholic. It is only honest to report that the evaluator of a family counseling program was going through a difficult divorce at the time of fieldwork.) Who funded the study and under what arrangements with the researcher? How did the researcher gain access to the study site? What prior knowledge did the researcher bring to the research topic and the study site?

There can be no definitive list of questions that must be addressed to establish investigator credibility. The principle is to report any personal and professional information that may have affected data collection, analysis, and interpretation either negatively or positively in the minds of users of the findings. For example, health status should be reported if it affected the researcher's stamina in the field. (Were you sick part of the time? The fieldwork for evaluation of an African health project was conducted over

three weeks during which time the evaluator had severe diarrhea. Did that affect the highly negative tone of the report? The evaluator said it didn't, but I'd want to have the issue out in the open to make my own judgment.) Background characteristics of the researcher (e.g., gender, age, race, ethnicity) may also be relevant to report in that such characteristics can affect how the researcher was received in the setting under study and related issues.

In preparing to interview farm families in Minnesota I began building up my tolerance for strong coffee a month before the fieldwork was scheduled. As ordinarily a non-coffee drinker, I knew my body would be jolted by 10–12 cups of coffee a day doing interviews in farm kitchens. These are matters of personal preparation, both mental and physical, that affect perceptions about the quality of the study. Preparation and training are reported as part of the study's methodology.

### *Researcher Training, Experience, and Preparation*

Every student who takes an introductory psychology or sociology course learns that human perception is highly selective. When looking at the same scene, design, or object, different people will see different things. What people "see" is highly dependent on their interests, biases, and backgrounds. Our culture tells us what to see; our early childhood socialization instructs us in how to look at the world; and our value systems tell us how to interpret what passes before our eyes. How, then, can one trust observational data? Or qualitative analysis?

In their popular text, *Evaluating Information: A Guide for Users of Social Science Research*, Katzer, Cook, and Crouch (1978) titled their chapter on observation "Seeing Is Not Believing." In that chapter they tell an oft-repeated story meant to demonstrate the problem with observational data:

Once at a scientific meeting, a man suddenly rushed into the midst of one of the sessions. He was being chased by another man with a revolver. They scuffled in plain view of the assembled researchers, a shot was fired, and they rushed out. About twenty seconds had elapsed. The chairperson of the session immediately asked all present to write down an account of what they had seen. The observers did not know that the ruckus had been planned, rehearsed, and photographed. Of the forty reports turned in, only one was less than 20-percent mistaken about the principal facts, and most were more than 40-percent mistaken. The event surely drew the undivided attention of the observers, was in full view at close range, and lasted only twenty seconds. But the observers could not observe all that happened. Some readers chuckled because the observers were researchers, but similar experiments have been reported numerous times. They are alike for all kinds of people. (pp. 21–22)

Research and experimentation on selective perception document the inadequacies of ordinary human observation and certainly cast doubt on the validity and reliability of observation as a major method of scientific inquiry. Yet there are two fundamental fallacies in generalizing from stories like the one about the inaccurate observations of researchers at the scientific meeting: (1) these researchers were not trained as qualitative observers and (2) they were not prepared to make observations at that particular moment in time. *Scientific inquiry using observational methods requires disciplined training and rigorous preparation.*

The simple fact that a person is equipped with functioning senses does not make that person a skilled observer. The fact that ordinary persons experiencing any particular situation will experience and perceive that situation differently does not mean that trained and prepared observers cannot report with accuracy, validity, and reliability the nature of that situation. Training includes learning how to write descriptively; practicing the disciplined recording of field notes; knowing how to separate detail from trivia in order to achieve the former without being overwhelmed by the latter; and using rigorous methods to validate observations. Training researchers to become astute and skilled observers is particularly difficult because so many people think that they are “natural” observers and, therefore, that they have very little to learn. Training to become a skilled observer is a no less rigorous process than the training necessary to become a skilled statistician. People don’t “naturally” know statistics—and people do not “naturally” know how to do systematic research observations. Both require training, practice, and preparation.

Careful preparation for making observations is as important as disciplined training. While I have invested considerable time and effort in becoming a trained observer, I am confident that, had I been present at the scientific meeting where the shooting scene occurred, my recorded observations would not have been significantly more accurate than those of my less trained colleagues. The reason is that I would not have been prepared to observe what occurred and, lacking that preparation, would have been seeing things through my ordinary participant’s eyes rather than my scientific observer’s eyes.

Preparation has mental, physical, intellectual, and psychological dimensions. Earlier I quoted Pasteur: “In the fields of observation, chance favors the prepared mind.” Part of preparing the mind is learning how to concentrate during the observation. Observation, for me, involves enormous energy and concentration. I have to “turn on” that concentration: “turn on” my scientific eyes and ears and taste, touch, and smell mechanisms. A scientific observer

cannot be expected to engage in systematic observation on the spur of the moment any more than a world-class boxer can be expected to defend his title spontaneously on a street corner or an Olympic runner can be asked to dash off at record speed because someone suddenly thinks it would be nice to test the runner's time. Athletes, artists, musicians, dancers, engineers, and scientists require training and mental preparation to do their best. Experiments and simulations that document the inaccuracy of spontaneous observations made by untrained and unprepared observers are no more indicative of the potential quality of observational methods than an amateur community talent show is indicative of what professional performers can do.

Two points are critical here. First, the folk wisdom about observation being nothing more than selective perception is true in the ordinary course of participating in day-to-day events. Second, the skilled observer is able to improve the accuracy, validity, and reliability of observations through intensive training and rigorous preparation.

Finally, let me add a few words about the special case of conducting evaluations, which adds additional criteria for assessing quality. A specialized profession of evaluation research has emerged with its own research base, practice wisdom, and standards of quality (Patton 1997). Being trained as a social scientist or researcher, or having an advanced degree, does not make one qualified, competent, or prepared to engage in the specialized craft of program evaluation. The notion that simply being an economist (or any other brand of social scientist) makes one qualified to conduct evaluations of the economic or other benefits of a program is one of the major contributors to the useless drivel that characterizes so many such evaluations. Of course, as a professional evaluator, one active in the profession and a member of the American Evaluation Association, my conclusions in this regard may be considered biased. I invite you, therefore, to undertake your own inquiry and reach your own conclusions by comparing evaluations conducted by professional health evaluators active in the evaluation profession in contrast to evaluations conducted by discipline-based health researchers with no special evaluation training or expertise. In that comparison, be sure to attend to the actual use and practicality of the resulting findings, for use and application are two of the primary standards that the profession has adopted for judging the quality of evaluations.

### *Researcher and Research Effects*

There are four ways in which the presence of the researcher, or the fact that an evaluation is taking place, can directly affect the findings of a study:

1. reactions of program participants and staff to the presence of the qualitative fieldworker;
2. changes in the fieldworker (the measuring instrument) during the course of the data collection or analysis, that is, instrumentation effects;
3. the predispositions, selective perceptions, and/or biases of the qualitative researcher; and
4. researcher incompetence (including lack of sufficient training or preparation).

The presence of a fieldworker can certainly make a difference in the setting under study. The fact that data are being collected may create a halo effect so that staff perform in an exemplary fashion and participants are motivated to "show off." Some emergent forms of program evaluation, especially "empowerment evaluation" and "intervention-reinforcing evaluation" (Patton 1997) turn this traditional threat to validity into an asset by designing data collection to enhance achievement of the desired program outcomes. For example, at the simplest level, the observation that "what gets measured gets done" suggests the power of data collection to affect outcomes attainment. A leadership program, for example, that includes in-depth interviewing and participant journal-writing as ongoing forms of evaluation data collection will find that participating in the interviewing and in writing reflectively have effects on participants and program outcomes. Likewise, a community-based AIDS awareness intervention can be enhanced by having community participants actively engaged in identifying and doing case studies of critical community incidents.

On the other hand, the presence of the fieldworker or evaluator may create so much tension and anxiety that performances are below par. Problems of reactivity are well documented in the anthropological literature. That is one of the prime reasons why qualitative methodologists advocate long-term observations that permit an initial period during which evaluators (observers) and the people in the setting being observed get a chance to get used to each other. Denzin's advice concerning the reactive effects of observers is, I think, applicable to the specific case of evaluator-observers:

It is axiomatic that observers must record what they perceive to be their own reactive effects. They may treat this reactivity as bad and attempt to avoid it (which is impossible), or they may accept the fact that they will have a reactive effect and attempt to use it to advantage. . . . The reactive effect will be measured by daily field notes, perhaps by interviews in which the problem is pointedly inquired about, and also in daily observations. (Denzin 1978: 200)

In brief, the qualitative researcher or evaluator has a responsibility to think about the problem, make a decision about how to handle it in the field, and then attempt to monitor observer effects. In so doing, it is worth noting that observer effects can be considerably overestimated. Lillian Weber, director of the Workshop Center for Open Education, City College School of Education, New York, once set me straight on this issue, and I pass her wisdom on to my colleagues. In doing observations of open classrooms, I was concerned that my presence, particularly the way kids flocked around me as soon as I entered the classroom, was distorting the situation to the point where it was impossible to do good observations. Lillian laughed and suggested to me that what I was experiencing was the way those classrooms actually were. She went on to note that this was common among visitors to schools; they were always concerned that the teacher, knowing visitors were coming, whipped the kids into shape for those visitors. She suggested that under the best of circumstances a teacher might get kids to move out of habitual patterns into some model mode of behavior for as much as 10 or 15 minutes, but that, habitual patterns being what they are, kids would rapidly revert to normal behaviors and whatever artificiality might have been introduced by the presence of the visitor would likely become apparent.

Qualitative researchers should strive neither to overestimate nor to underestimate their effects but to take seriously their responsibility to describe and study what those effects are.

The second concern about evaluator effects arises from the possibility that the evaluator changes during the course of the evaluation. One of the ways this sometimes happens in anthropological research is when a participant observer "goes native" and is absorbed into the local culture. The epitome of this in a shorter-term observation is the story of the observers who became converted to Christianity while observing a Billy Graham crusade (Lang and Lang 1960). In the health area this is represented by the classic dilemma of the anthropologist in a village of poor people who struggles with whether to offer basic nutrition counsel to save a baby's life. Program evaluators using participant observation sometimes become personally involved with program participants or staff in ways that affect their own attitudes and behaviors. The consensus of advice on how to deal with the problem of changes in the observers as a result of involvement in research is the same as advice about how to deal with the reactive effects created by the presence of observers: pay attention and record these changes, and stay sensitive to shifts in one's own perspective by systematically recording one's own perspective at points during fieldwork.

The third concern about evaluator effects has to do with the extent to which the predispositions or biases of the evaluator may affect data analysis and interpretations. This issue involves a certain amount of paradox because, on the one hand, rigorous data collection and analytical procedures, like triangulation, are aimed at substantiating the validity of the data; on other hand, the interpretive and social construction underpinnings of the phenomenological paradigm mean that data from and about humans inevitably represent some degree of perspective rather than absolute truth. Getting close enough to the situation observed to experience it firsthand means that researchers can learn from their experiences, thereby generating personal insights, but that closeness makes their objectivity suspect.

This paradox is addressed in part through the notion of "emphatic neutrality," a stance in which the researcher or evaluator is perceived as caring about and being interested in the people under study, but as being neutral about the findings. House suggests that the evaluation researcher must be seen as impartial:

The evaluator must be seen as caring, as interested, as responsive to the relevant arguments. He must be impartial rather than simply objective. The impartiality of the evaluator must be seen as that of an actor in events, one who is responsive to the appropriate arguments but in whom the contending forces are balanced rather than nonexistent. The evaluator must be seen as not having previously decided in favor of one position or the other. (House 1977: 45-46)

Neutrality and impartiality are not easy stances to achieve. Denzin (1989) cites a number of scholars who have concluded, as he does, that all researchers bring their own preconceptions and interpretations to the problem being studied, regardless of the methods used:

All researchers take sides, or are partisans for one point of view or another. Value-free interpretive research is impossible. This is the case because every researcher brings preconceptions and interpretations to the problem being studied. The term "hermeneutical circle or situation" refers to this basic fact of research. All scholars are caught in the circle of interpretation. They can never be free of the hermeneutical situation. This means that scholars must state beforehand their prior interpretations of the phenomenon being investigated. Unless these meanings and values are clarified, their effects on subsequent interpretations remain clouded and often misunderstood. (p. 23)

Debate about the research value of qualitative methods means that researchers must make their own peace with how they are going to describe what they do. The meaning and connotations of words like *objectivity*, *subjectivity*, *neutrality*, and *impartiality* will have to be worked out with particular



audiences in mind. Essentially, these are all concerns about the extent to which the qualitative researcher can be trusted; that is, trustworthiness is one dimension of perceived methodological rigor (Lincoln and Guba 1985). For better or worse, the trustworthiness of the data is tied directly to the trustworthiness of the researcher who collects and analyzes the data.

The final issue affecting evaluator effects is that of competence. Competence is demonstrated by using the verification and validation procedures necessary to establish the quality of analysis, and it is demonstrated by building a track record of fairness and responsibility. Competence involves neither overpromising nor underproducing in qualitative research.

### *Intellectual Rigor*

The thread that runs through this discussion of researcher credibility is the importance of intellectual rigor and professional integrity. There are no simple formulas or clear-cut rules to direct the performance of a credible, high-quality analysis. The task is to do one's best to make sense out of things. A qualitative analyst returns to the data over and over again to see if the constructs, categories, explanations, and interpretations make sense, if they really reflect the nature of the phenomena. Creativity, intellectual rigor, perseverance, insight—these are the intangibles that go beyond the routine application of scientific procedures. As Nobel prize winning physicist Percy Bridgman put it: "There is no scientific method as such, but the vital feature of a scientist's procedure has been merely to do his utmost with his mind, *no holds barred* (quoted in Mills 1961: 58, emphasis in the original).

## THE PARADIGMS DEBATE AND CREDIBILITY

Having presented techniques for enhancing the quality of qualitative analysis, such as triangulation, and having discussed ways of addressing the perceived trustworthiness of the investigator, the third and final credibility issue involves philosophical beliefs about the rationale for, appreciation of, and worthwhileness of naturalistic inquiry, qualitative methods, inductive analysis, and holistic thinking—all core paradigm themes. Much of the controversy about qualitative methods stems from the long-standing debate in science over how best to study and understand the world. This debate sometimes takes the form of qualitative versus quantitative methods or logical positivism versus phenomenology. The debate is rooted in philosophical differences about the

nature of reality. Several sources provide a detailed discussion of what has come to be called "the paradigms debate," a paradigm being a particular worldview (see Patton 1997). The point here is to remind health researchers about the intensity of the debate. It is important to be aware that both scientists and nonscientists often hold strong opinions about what constitutes credible evidence. Given the potentially controversial nature of methods decisions, researchers using qualitative methods need to be prepared to explain and defend the value and appropriateness of qualitative approaches. The sections that follow will briefly discuss the most common concerns.

### *Methodological Respectability*

As Thomas Cook, one of evaluation's luminaries, pronounced in his keynote address to the 1995 International Evaluation Conference in Vancouver, "Qualitative researchers have won the qualitative-quantitative debate."

Won in what sense? Won acceptance.

The validity of experimental methods and quantitative measurement, appropriately used, was never in doubt. Qualitative methods are now ascending to a level of parallel respectability, particularly among evaluation researchers. A consensus is emerging in the evaluation profession that researchers and evaluators need to know and use a variety of methods in order to be responsive to the nuances of particular empirical questions and the idiosyncrasies of specific stakeholder needs. The issue is the appropriateness of methods for a specific evaluation research purpose and question, not adherence to some absolute orthodoxy that declares one or the other approach to be inherently preferred. The problem is that this ideal, characterizing evaluation researchers as situationally responsive, methodologically flexible, and sophisticated in using a variety of methods, runs headlong into the realities of the evaluation world. Those realities include limited resources, political considerations of expediency, and the narrowness of disciplinary training available to most graduate students—training that imbues them with varying degrees of methodological prejudice.

A paradigm is a worldview built on implicit assumptions, accepted definitions, comfortable habits, values defended as truths, and beliefs projected as reality. As such, paradigms are deeply embedded in the socialization of adherents and practitioners: paradigms tell them what is important, legitimate, and reasonable. Paradigms are also normative, telling the practitioner what to do without the necessity of long existential or epistemological consideration.

But it is this aspect of paradigms that constitutes both their strength and their weakness—their strength in that it makes action possible, their weakness in that the very reason for action is hidden in the unquestioned assumptions of the paradigm.

### *Beyond the Numbers Game*

Thomas H. Kuhn is a philosopher of science who has extensively studied the value systems of scientists. Kuhn (1970: 184–85) has observed that “the most deeply held values concern predictions.” He goes on to observe that “quantitative predictions are preferable to qualitative ones.” The methodological status hierarchy in science ranks “hard data” above “soft data” where “hardness” refers to the precision of statistics. Qualitative data, then, carry the stigma of being “soft.”

I’ve already argued that among leading methodologists in fields like evaluation research, this old bias has diminished or even disappeared. But elsewhere it prevails. How can one deal with a lingering bias against qualitative methods?

The starting point is understanding and being able to communicate the particular strengths of qualitative methods and the kinds of empirical questions for which qualitative data are especially appropriate. It is also helpful to understand the special seductiveness of numbers in modern society. Numbers convey a sense of precision and accuracy even if the measurements that yielded the numbers are relatively unreliable, invalid, and meaningless. The point, however, is not to be anti-numbers. The point is to be *pro-meaningfulness*. Moreover, as noted in discussing the value of methods triangulation, the issue need not be quantitative *versus* qualitative methods, but rather how to combine the strengths of each in a multimethods approach to research and evaluation. Qualitative methods are not weaker or softer than quantitative approaches; qualitative methods are *different*.

## THE CREDIBILITY ISSUE IN SUMMARY

This overview has examined ways of enhancing the quality and credibility of qualitative analysis by dealing with three distinct but related inquiry concerns: rigorous techniques and methods for gathering and analyzing qualitative data; the credibility, competence, and perceived trustworthiness of the qualitative researcher; and philosophical beliefs or paradigm-based preferences such as objectivity versus subjectivity and generalizations versus extrapolations. In

the early evaluation literature, the debate between qualitative and quantitative methodologists was often strident. In recent years it has softened. A consensus has gradually emerged that the important challenge is to match methods appropriately to empirical questions and issues, and not to universally advocate any single methods approach for all problems.

In a diverse world, one aspect of diversity is methodological. From time to time regulations surface in various federal and state agencies that prescribe universal, standardized evaluation measures and methods for all programs funded by those agencies. I oppose all such regulations in the belief that local program processes are too diverse and client outcomes too complex to be fairly represented nationwide, or even statewide, by some narrow set of prescribed measures and methods—regardless whether the mandate be for quantitative or for qualitative approaches. When methods decisions are based on some universal, political mandate rather than on situational merit, research offers no challenge, requires no subtlety, presents no risk, and allows for no accomplishment.

## REFERENCES

- Denzin, N. K. 1989. *Interpretive Interactionism*. Newbury Park, CA: Sage.
- . 1978. *The Research Act: A Theoretical Introduction to Sociological Methods*. New York: McGraw-Hill.
- Fielding, N. J., and J. L. Fielding. 1986. *Linking Data*. Qualitative Methods Series No. 4. Newbury Park, CA: Sage Publications.
- House, E. 1977. *The Logic of Evaluative Argument*. CSE Monograph Series in Evaluation no. 7. Los Angeles: University of California, Los Angeles, Center for the Study of Evaluation.
- Katzer, J., K. Cook, and W. Crouch. 1978. *Evaluating Information: A Guide for Users of Social Science Research*. Reading, MA: Addison-Wesley.
- Kuhn, T. 1970. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.
- Lang, K., and G. E. Lang. 1960. "Decisions for Christ: Billy Graham in New York City." In *Identity and Anxiety*, edited by M. Stein, A. Vidich, and D. White. New York: Free Press.
- Lincoln, Y., and E. G. Guba. 1985. *Naturalistic Inquiry*. Newbury Park, CA: Sage.
- Mills, C. W. 1961. *The Sociological Imagination*. New York: Oxford University Press.
- Patton, M. Q. 1999. *Grand Canyon Celebration: A Father-Son Journey of Discovery*. Amherst, NY: Prometheus Books.
- . 1997. *Utilization-Focused Evaluation: The New Century Text*. Thousand Oaks, Ca: Sage Publications.
- . 1990. *Qualitative Evaluation and Research Methods*. Thousand Oaks, CA: Sage Publications.